DETERMINATION OF RELIABLE REFERENCE GENES FOR REVERSE TRANSCRIPTION QUANTITATIVE REAL-TIME PCR FROM OIL PALM TRANSCRIPTOMES

NADZIRAH AMIRUDDIN1; PEK-LAN CHAN1; KUANG-LIM CHAN1; PEI-WEN ONG1; PRISCILLA ELIZABETH MORRIS1; MEILINA ONG-ABDULLAH1; SUBHI SITI MASURA1 and ENG-TI LESLIE LOW1*

ABSTRACT
A set of reliable reference genes is essential for accurate quantification and interpretation of gene expression data using reverse transcription quantitative real-time polymerase chain reaction (RT-qPCR). In this study, Roche-454 RNA-seq reads from 27 libraries of various oil palm tissues were systematically analysed to identify a set of potential reference genes. Eleven candidate reference genes were identified from the transcriptome data. These genes, together with three oil palm reference genes previously identified for tissue culture samples (PD000380, PD00569, pOPEA01332) and five classical housekeeping genes [glyceraldehyde-3-phosphate dehydrogenase (GAPDH), NAD5, TUBULIN, UBIQUITIN, ACTIN] were analysed across samples collected from various tissues from mature oil palm (leaf, root, endosperm, mesocarp, female flowers) and stages of tissue culture (non-embryogenic callus, embryogenic callus, polyembryoids, and shoots from polyembryoids). The expression levels of these genes were compared and evaluated using geNorm. Three genes (EgREF_5, EgREF_7 and EgREF_11) were found to be appropriate reference genes for normalising gene expression data from both mature plant and tissue culture samples.

Keywords: housekeeping genes, normalisation, RT-qPCR, transcriptome.

Received: 21 May 2021; Accepted: 2 November 2021; Published online: 17 January 2022.

INTRODUCTION
Gene expression analysis or the study of transcript abundance is essential in molecular biology research, especially in understanding the role of gene expression patterns in different biological processes (Dussert et al., 2013; Kong et al., 2021; Tranbarger et al., 2011). It furthers our understanding and provides insights on the genetic and molecular mechanisms underlying developmental and cellular processes.

Reverse transcription quantitative real-time polymerase chain reaction (RT-qPCR) is one of the most powerful and sensitive techniques to quantify gene expression levels and has been recognised as a key driver of gene expression analysis in numerous molecular biology applications (Kozera and Rapacz, 2013; VanGuilder et al., 2008). It is still considered the method of choice to validate high-throughput gene expression data (Everaert et al., 2017). This technique is sensitive enough to detect gene expression changes, for even low transcript levels (Bustin et al., 2005; Nolan et al., 2006).

Nevertheless, it is important to note that substantial experimental variability, such as initial material quality, different inhibitors in samples, primer design and reverse transcription efficiencies should be taken into account to accurately quantify
gene expression (Ginzinger, 2002; Mahoney et al., 2004). Technical variability could be added to the data by random pipetting errors (Bustin and Nolan, 2004; Fleige and Pfaffl, 2006). These factors can cause unreliability in the quantifications of gene transcripts. Despite being considered highly sensitive, accurate, and reproducible (Kubista et al., 2006), this approach needs normalisation of the expression data to reduce the effects of variability in experimental data. Selection of an appropriate normalisation strategy is critical to acquire biologically meaningful data. Commonly, the level of gene expression is normalised by comparing messenger ribonucleic acid (mRNA) levels of target genes to endogenous controls, known as reference genes.

Reliable reference genes are required to interpret quantification data from RT-qPCR and compensate for any differences in the studied tissues or cells. An ideal reference gene is expressed at a constant level across various conditions and unaffected by experimental parameters (Guénin et al., 2009; Schmittgen and Zakrzajsek, 2000; Thellin et al., 1999; Zhu et al., 2013). In addition, they should not be co-regulated with the target gene but must be expressed in abundance with minimal variability (Radonic et al., 2004). The reference and target genes should also have similar ranges of expression in the samples to be analysed (Cappelli et al., 2008). The use of one or more reference genes is preferred for optimum normalisation (Bustin et al., 2009; Chandna et al., 2012; Guénin et al., 2009; Vandesompele et al., 2002; 2009). However, expression levels of reference genes can vary in response to changes in experimental conditions and/or tissue types, and that using unstable reference genes in the relative quantification of gene expression will lead to biases and inappropriate biological data interpretation (Artico et al., 2010; Czechowski et al., 2005; Dheda et al., 2005; Li et al., 2020; Thellin et al., 1999).

The expression of classical housekeeping genes such as glyceraldehyde-3-phosphate dehydrogenase (GAPDH), actin, tubulin and 18S ribosomal RNA were presumed to have constant levels of expression, being involved in basal cell metabolism, cytoskeleton or intracellular functions. These genes are widely used as reference genes. However, their expression levels have been demonstrated to vary under different experimental conditions (Barsalobres-Cavallari et al., 2009; Exposito-Rodriguez et al., 2008; Jian et al., 2008; Karuppaiya et al., 2017; Li et al., 2020; Mallona et al., 2010; Qi et al., 2010; Reid et al., 2006; Tang et al., 2021).

Meta-analysis of transcript data has been shown to be an efficient method to survey for novel stably expressed genes from high-throughput technologies as an alternative to mine for reference genes. For example, microarray datasets have enabled the identification of novel reference genes from a variety of plants, including Arabidopsis thaliana, Eucalyptus, soybean and rice, as well as in human samples and cynomolgus monkeys (Chan et al., 2014; Chang et al., 2011; Cheng et al., 2011; Czechowski et al., 2005; de Oliveira et al., 2012; Garrido et al., 2020; Libault et al., 2008; Narsaï et al., 2010; Park et al., 2013).

Another high-throughput analysis method which uses deep-sequencing technologies known as RNA-seq, has provided an additional resource to microarrays. It is widely used for quantitative mRNA expression studies and sensitive enough to detect very low transcript expressions and their isoforms (Marioni et al., 2008; Wang et al., 2009). Furthermore, it also has the ability to identify novel transcripts and splice variants (Trapnell et al., 2010). The technique is fast and generates replicated data with minimal variations (Marioni et al., 2008; Mortazavi et al., 2008; Nagalakshmi et al., 2008; Wang et al., 2009). Thus, RNA-seq is a feasible whole-transcriptome method for mining of stably expressed genes and the identification of novel reference genes for RT-qPCR normalisation. In recent findings, combinations of suitable reference genes for abiotic stresses in potato (Tang et al., 2017) and Arabidopsis pintula (Jin et al., 2019) were identified from transcriptome data.

RNA-seq derived transcriptome data has been widely explored in a number of oil palm studies (Beulé et al., 2011; Bourgis et al., 2011; Dussert et al., 2013; Kong et al., 2021; Shearman et al., 2013). In oil palm, RNA-seq data was generated in studies on lipid accumulation (Bourgis et al., 2011) and carbon partitioning (Bourgis et al., 2011) in mesocarp, response to Ganoderma boninense infection (Tee et al., 2013), normal or mantled flowers and fruits (Shearman et al., 2013), and phosphorus starvation in roots (Kong et al., 2021). More research was also conducted with the availability and increasing volume of African oil palm sequence data (Adam et al., 2007; Bourgis et al., 2011; Chan et al., 2017; Jaligot et al., 2011; Low et al., 2008; 2014; Singh et al., 2013). These datasets as a whole provide an invaluable pool of information that can be utilised to identify stably expressed genes that would greatly contribute to accurate and reliable quantification of RT-qPCR data. In oil palm, a number of reference genes for normalisation of gene expression has been identified inspecific developmental stages, such as tissue culture (Chan et al., 2014), stress-treated samples (Xia et al., 2014), diverse sets of biological samples (vegetative and reproductive tissues, and developmental stages of mesocarp tissues) (Yeap et al., 2014), as well as from young plantlets (Muhammad Afiq et al., 2019).

Nevertheless, improving the collection of reference genes would certainly help expedite validation of newly discovered genes involved in various tissues and developmental stages.

In this study, Roche-454 RNA-seq transcriptome libraries comprising 27 oil palm tissue samples were utilised to identify a set of novel reference
genes for normalisation of RT-qPCR expression data in oil palm. Nineteen candidate reference genes including three oil palm reference genes (PD00380, PD00569, pOP-EA01332) previously identified for tissue culture samples (Chan et al., 2014), five classical housekeeping genes (GAPDH, NAD5, TUBULIN, UBIQUITIN, ACTIN), and 11 genes from the transcriptome data were evaluated. Statistical analysis using geNorm identified the three most stable reference genes in combination, EgREF_5, EgREF_7 and EgREF_11. This result demonstrates that stably expressed reference genes could be identified from mining and utilisation of transcriptome datasets.

MATERIALS AND METHODS

Plant Materials

Leaf (spear and mature), endosperm [10 weeks after anthesis (WAA) and 15WAA], mesocarp (10WAA and 15WAA) and female inflorescence (frond number F13: 8.5 cm in length and F17: 20 cm) were sampled from Malaysian Palm Oil Board (MPOB) Research Station, Kluang, Johor, Malaysia. White root from seedlings in polybags, primary root (10 months nursery palms) and lateral root (10 months nursery palms) were sampled from the oil palm nursery in MPOB, Bangi, Selangor, Malaysia. Embryogenic callus, non-embryogenic callus, polyembryoids, and shoots from polyembryoids were collected from the tissue culture laboratory at the Advanced Biotechnology and Breeding Centre, MPOB, Bangi, Selangor, Malaysia. All samples were frozen in liquid nitrogen and stored at -80ºC prior to RNA extraction.

Ribonucleic Acid (RNA) Extraction, Purification and Quality Assessment

Total RNA was extracted according to the method by Ong et al. (2019). The total RNA was purified using RNeasy Mini Kit with on-column RNase free DNase I treatment according to the manufacturer’s instructions (Qiagen, USA, Valencia, CA, USA). The concentration and purity of total RNA were determined using Nanodrop ND-1000 UV-Vis Spectrophotometer (Thermo Fisher Scientific Inc.), and the integrity was assessed by electrophoretic fractionation on an Agilent 2100 Bioanalyzer using RNA 6000 Nano LabChip (Agilent Technologies, CA, USA).

Identification of Candidate Reference Genes in Oil Palm Transcriptome Data Sets

Candidate reference genes in oil palm were identified by performing differential expression analysis of Roche-454 RNA-seq reads from 27 libraries of various oil palm tissues from PRJNA201497 (leaf, root, seedling white root, pollen, mesocarp, endosperm) (Singh et al., 2013) and PRJNA345530 [pith (one day after anthesis, DAA), sepal (1DAA), fruit (1DAA), spikelet (1DAA), stalk (1DAA)] (Chan et al., 2017). Reads of each tissue were mapped to the published oil palm EG5 reference genome (Singh et al., 2013) using TopHat2 (Kim et al., 2013). Gene expression data was generated by cuffdiff (part of TopHat2 package) and normalised using the Fragments Per Kilobase per Million Reads method (FPKM). The average expression of genes expressed in each tissue was calculated and genes with a minimum FPKM of 40 in all tissues were selected as putative stably expressed genes. These identified genes were compared to the protein database using BLASTX (https://blast.ncbi.nlm.nih.gov) with default parameter I e⁻⁵.

Primer Design and Efficiency Test

Primer pairs for candidate reference genes were designed using Primer3Plus software (http://www.bioinformatics.nl/primer3plus) (Untergasser et al., 2007) with melting temperatures between 60°C and 67°C, primer lengths 20-27 bases, GC content 40%-60%, and amplicon lengths of 100-150 bp (Chan et al., 2014). The transcript sequences of candidate reference genes used for primer design were aligned to the oil palm EG5 genome sequence to predict intron positions using Exonerate program (Slater and Birney, 2005). Primer pairs were designed on either different exons or spanning exon-exon junctions of the complementary deoxyribonucleic acid (cDNA) (Hu et al., 2009) to avoid co-amplification of the genes from genomic DNA. Primer search against the oil palm genome was performed to check the specificity of each designed primer. The HPLC-purified primers were purchased from Bio Basic Canada Inc. All primer pairs were tested for polymerase chain reaction (PCR) amplification efficiencies (Ex) and to check for the specificity of the amplicon.

Reverse Transcription Quantitative Real-time PCR (RT-qPCR)

First-strand cDNA synthesis was carried out using 2 μg of total RNA using the High-capacity cDNA Reverse-Transcription Kit according to the manufacturer’s instructions (Applied Biosystems, Foster City, CA, USA). The synthesised cDNA strands were used as templates in a SYBR Green based RT-qPCR using the Eppendorf Mastercycler® ep realplex (Eppendorf, Germany). RT-qPCR was performed according to the method described by Chan et al. (2014). A ‘no reverse transcriptase’ (NRT) control and a ‘non-template control’ (NTC) were assigned as negative controls.
Data Analysis

Cycle threshold (Ct) values were retrieved using Realplex software version 2.2 (Eppendorf, Germany). Results were imported into Microsoft Excel and data analysis was carried out by calculating the average Ct values for three replicates. Subsequently, the results were transformed into expression quantities using the method described by Vandesompele et al. (2002), Ex^(minCt - sampleCt), (Ex = 10^[slope×(Ct-mean)] × 100%, slope = slope of linear regression). The most stable reference genes across all samples were selected based on geNorm v3.4 using log-transformed data as input (Vandesompele et al., 2002).

RESULTS

Selection of Candidate Reference Genes and Expression Analysis

A total of 11 putative stably expressed genes were identified through differential gene expression analysis of 27 oil palm transcriptome libraries. These genes are involved in various functions, such as DNA binding protein, macrophage migration inhibitory factor family protein, histone, deleted in split hand/split foot protein, MKI67 FHA domain-interacting nucleolar phosphoprotein-like, NADH-ubiquinone oxidoreductase 13 kDa-B subunit, ribosomal protein, OB-fold nucleic acid binding domain-containing protein and ubiquitin. Of all the 11 genes, EgREF_10, which is a ubiquitin family protein (Table 1), showed the highest level of variation, with the highest expression in pollen and floret after anthesis (Figure 1). The difference between the highest and lowest was ~2.2 fold difference in expression while the average difference for the other genes was ~1.6 fold. Primer pairs were designed for the 11 genes from the transcriptome data, five classical housekeeping genes and three oil palm tissue culture samples reference genes (Chan et al., 2014). Efficiency of designed primer pairs was evaluated by the presence of a single peak in the melting curve obtained after 40 cycles of amplification. Only primer pairs which showed a single amplified product were selected for further studies. Thirteen candidate genes were thus selected for further analysis (Table 2). The Ex of the selected primers ranged from 82% to 100%, and correlation coefficient (R²) ranged from 0.979 to 0.999. Subsequently, SYBR Green detection-based RT-qPCR assay was carried out for transcript profiling of these genes. As shown in Figure 2, the genes have an average Ct value of 22.79. Most of the genes have expression values between 20 and 24. UBIQUITIN was the most abundant gene of the set (mean Ct: PD00380 = 25.58, pOP-EA01332 = 25.98). PD00380 showed highest level of Ct variation, with a range between 33.91 and 22.23, of which the difference between 25th and 75th percentile is ~5 Ct.

Selection of Potential Reference Genes for Oil Palm

GeNorm analysis executed using a Visual Basic Application in Microsoft Excel (Vandesompele et al., 2002) was used to select the most stably expressed reference gene across various oil palm tissues. Gene expression stability measure (M) of each reference gene was calculated using the relative expression values for each cDNA sample as input for the geNorm algorithm. The threshold proposed for stably expressed genes was M<0.5 (Vandesompele et al., 2002). Genes with the lowest M value are deemed to have the most stable expression. Two most stably expressed genes were obtained by eliminating the least stable gene in a stepwise manner. The algorithm ranked the potential reference genes based on their expression stability (Figure 3). EgREF_5 and EgREF_7 were the most stable genes having M~0.5, while other genes (EgREF_11, pOP-EA01332, PD00569, EgREF_1, ACTIN, GAPDH, EgREF_2, EgREF_5, UBIQUITIN) had an M value between 0.5 and 1. The least stable genes, TUBULIN and PD00380, had M values of more than 1.

GeNorm was also used to determine the optimal number of reference genes required for accurate and reliable normalisation of expression data in the tested sample sets. Parameter V, defined as the pairwise variation (Vn/Vn-1) was calculated by geNorm between the two sequential normalisation factors (NF) of n and n + 1 genes of consecutively ranked reference genes. The purpose is to measure the effect of adding further reference genes on the normalisation factor. A cut-off threshold parameter V of 0.15 was recommended by Vandesompele et al. (2002), below which the addition of more reference genes is not required. As shown in Figure 4, the analysis showed that V3/V4 was 0.129, suggesting that the optimum number of reference genes is three. Hence, the best combination of reference genes is EgREF_5, EgREF_7 and EgREF_11. The data also indicates that the new reference genes have greater expression stability than the conventionally used housekeeping genes, and can therefore provide more reliable normalised expression data.

DISCUSSION

RT-qPCR is one of the most sensitive tools that is commonly used to achieve rapid and reliable quantification of gene expression levels. However,
TABLE 1. SELECTED CANDIDATE REFERENCE GENES FROM TRANSCRIPTOME DATA

<table>
<thead>
<tr>
<th>Query</th>
<th>Acc. num.</th>
<th>Annotation</th>
<th>E-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>EgREF_1</td>
<td>NP_001147321</td>
<td>DNA binding protein (Zea mays)</td>
<td>2.00E-11</td>
</tr>
<tr>
<td>EgREF_2</td>
<td>NP_195785</td>
<td>Macrophage migration inhibitory factor family protein (Arabidopsis thaliana)</td>
<td>2.00E-51</td>
</tr>
<tr>
<td>EgREF_3</td>
<td>XP_002319269</td>
<td>Histone 2 (Populus trichocarpa)</td>
<td>1.00E-43</td>
</tr>
<tr>
<td>EgREF_4</td>
<td>XP_002870907</td>
<td>Macrophage migration inhibitory factor family protein (Arabidopsis lyrata subsp. Lyrate)</td>
<td>3.00E-52</td>
</tr>
<tr>
<td>EgREF_5</td>
<td>NM_001153724</td>
<td>Deleted in split hand/split foot protein (Zea mays)</td>
<td>1.00E-26</td>
</tr>
<tr>
<td>EgREF_6</td>
<td>NP_001241759</td>
<td>MKI67 FHA domain-interacting nucleolar phosphoprotein-like (Zea mays)</td>
<td>2.00E-66</td>
</tr>
<tr>
<td>EgREF_7</td>
<td>NP_001151372</td>
<td>NADH-ubiquinone oxidoreductase 13 kDa-B subunit (Zea mays)</td>
<td>9.00E-64</td>
</tr>
<tr>
<td>EgREF_8</td>
<td>NP_001146947</td>
<td>60S ribosomal protein L27 (Zea mays)</td>
<td>8.00E-59</td>
</tr>
<tr>
<td>EgREF_9</td>
<td>XP_178531</td>
<td>OB-fold nucleic acid binding domain-containing protein (Arabidopsis thaliana)</td>
<td>1.00E-35</td>
</tr>
<tr>
<td>EgREF_10</td>
<td>XP_00263802</td>
<td>Ubiquitin family protein (Arabidopsis lyrata subsp. Lyrate)</td>
<td>1.00E-33</td>
</tr>
<tr>
<td>EgREF_11</td>
<td>NP_001152653</td>
<td>40S ribosomal protein S14 (Zea mays)</td>
<td>2.00E-63</td>
</tr>
</tbody>
</table>

Figure 1. Expression of putative stably expressed genes in various oil palm tissues based on transcriptome data. (a) spear leaf (dura), spear leaf (pisifera), F17 leaf (pisifera), root (dura), root (pisifera), seedling white root, pollen (female fertile), pollen (female sterile) (b) mesocarp 10WAA, mesocarp 15WAA, endosperm 10WAA, endosperm 15WAA, fruit, pith, sepal, spikelet, stalk, and (c) floret dura after anthesis, floret tenera after anthesis, female flower abnormal (2 cm), female flower abnormal (19 cm), female flower normal (2.5 cm), female flower normal (20 cm), shoot apex normal, shoot apex abnormal.
### Table 2. Primers and Amplicon Characteristics of Selected Reference Genes

<table>
<thead>
<tr>
<th>Gene abbreviation</th>
<th>Gene description</th>
<th>GenBank ID</th>
<th>Primer sequences (F/R) (5'-3')</th>
<th>Amplicon length (bp)</th>
<th>Annealing temperature (°C)</th>
<th>Amplification efficiency (%)</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>EgREF_3</td>
<td>Histone 2</td>
<td>Pr032825864</td>
<td>ATTTCTCAAGGCGGCAAAGTACG GCTTCGAGCAATCTCTAGTT</td>
<td>150</td>
<td>60</td>
<td>85</td>
<td>0.996</td>
</tr>
<tr>
<td>EgREF_4</td>
<td>Macrophage migration inhibitory factor family protein</td>
<td>Pr032825863</td>
<td>CCAAGACTGTCGCGAACGTCATA GCCTCAATGAAAAACCAATTGCC</td>
<td>131</td>
<td>60</td>
<td>92</td>
<td>0.998</td>
</tr>
<tr>
<td>EgREF_5</td>
<td>Deleted in split hand/split foot protein 1</td>
<td>Pr032825865</td>
<td>AGGACGCGGAGATCGACCTCTTT ATCACTTCCCACCTGCGTGA</td>
<td>116</td>
<td>60</td>
<td>87</td>
<td>0.997</td>
</tr>
<tr>
<td>EgREF_7</td>
<td>NADH-ubiquinone oxidoreductase 13 kDa-B subunit</td>
<td>Pr032825866</td>
<td>GCCAAGATGATGCAATGGGACCC GTCGGTGTGGGGAACGTCCTT</td>
<td>103</td>
<td>60</td>
<td>82</td>
<td>0.998</td>
</tr>
<tr>
<td>EgREF_10</td>
<td>Ubiquitin family protein</td>
<td>Pr032825867</td>
<td>ACAAGATCCCCCATCAGAAGTGGT GCTGATTCTGATCAACCACCC AACC</td>
<td>135</td>
<td>60</td>
<td>83</td>
<td>0.997</td>
</tr>
<tr>
<td>EgREF_11</td>
<td>40S ribosomal protein S17</td>
<td>Pr032825868</td>
<td>TCTCCCTCAAGGCTGAGGGGAG GACATTCCTGGACGTCTCTGCTGC</td>
<td>111</td>
<td>60</td>
<td>87</td>
<td>0.995</td>
</tr>
<tr>
<td>PD00380</td>
<td>Predicted 40S ribosomal protein S27-1</td>
<td>YE397675</td>
<td>GATGTTCTTGGGACGTAATTTGA TCACATCCCAGAAGATGATTTGC</td>
<td>113</td>
<td>60</td>
<td>87</td>
<td>0.995</td>
</tr>
<tr>
<td>PD00569</td>
<td>Manganese superoxide dismutase</td>
<td>EL682210</td>
<td>ACCACACAGACGATACAGAAA GATATGACCTCCGCCAATGAAC</td>
<td>129</td>
<td>63</td>
<td>87</td>
<td>0.999</td>
</tr>
<tr>
<td>pOP-EA01332</td>
<td>Predicted protein IFH-1 like</td>
<td>YE406625</td>
<td>AAGGACGAAATGCGAAGTACAG ACGATAGCTCCAGCAGA AGTT</td>
<td>111</td>
<td>63</td>
<td>100</td>
<td>0.979</td>
</tr>
<tr>
<td>GAPDH</td>
<td>Glyceraldehyde 3-phosphate dehydrogenase</td>
<td>DQ267444</td>
<td>GATCGAGAAATCCAGACGTATG GTCACATCAAAGTCCGCTGGCA</td>
<td>124</td>
<td>63</td>
<td>87</td>
<td>0.979</td>
</tr>
<tr>
<td>TUBULIN</td>
<td>Alpha-tubulin 1</td>
<td>EL685625</td>
<td>CATGCCTTGTCGCTTGTAATGC CATGCCTTGTCGCTTGTAATGC</td>
<td>109</td>
<td>63</td>
<td>95</td>
<td>0.999</td>
</tr>
<tr>
<td>UBIQUITIN</td>
<td>Polyubiquitin</td>
<td>EL689143</td>
<td>CCAGGCACATCTTCACGATG GGGGATGCCCCTTCCCTTAC</td>
<td>130</td>
<td>63</td>
<td>93</td>
<td>0.998</td>
</tr>
<tr>
<td>ACTIN</td>
<td>Actin</td>
<td>AX750991</td>
<td>GTCCTGACTGTAGCAGCGACGAA GAAATCCACATCTGGGAACGGT</td>
<td>147</td>
<td>60</td>
<td>83</td>
<td>0.999</td>
</tr>
</tbody>
</table>

Figure 2. Mean Ct values of candidate reference genes across various oil palm tissues. The range of Ct values was exhibited in boxplot. The box indicates the 25th and 75th percentiles. Horizontal line inside the box is the median. Whiskers below and above the box represent minimum and maximum values. The 15 oil palm tissues tested are spear leaf, mature leaf, white root, lateral root, primary root, 10WAA kernel, 15WAA kernel, 10WAA mesocarp, 15WAA mesocarp, F17 inflorescence (20 cm), F13 inflorescence (8.5 cm), non-embryogenic callus, embryogenic callus, polyembryoids and shoots from the polyembryoids.
Normalisation is a critical step to compensate for technical variability caused by various steps of the experimental procedures which can affect interpretation of data from RT-qPCR. For this reason, it has become common practice to perform relative quantification by normalising RT-qPCR data to reference genes, as compared to absolute this approach is potentially affected by the quantity and quality of initial materials, first strand cDNA synthesis efficiency, primer performance and statistical analysis methods chosen (Maroufi et al., 2010). Thus, there is a need to normalise raw expression data with stably expressed reference genes for accurate and reliable results.

Normalisation is a critical step to compensate for technical variability caused by various steps of the experimental procedures which can affect interpretation of data from RT-qPCR. For this reason, it has become common practice to perform relative quantification by normalising RT-qPCR data to reference genes, as compared to absolute...
quantification in which the input copy number is determined by relating the PCR signal to a standard curve for the particular genes of interest. It is also essential to understand that measurement of gene expression patterns may vary under different experimental conditions. Therefore, validation of reference genes should be done according to the panel that is specific for the chosen experimental conditions and tissue types under which the target gene is studied (Ruan and Lai, 2007; Selvey et al., 2001; Song et al., 2020; Suzuki et al., 2000; Thellin et al., 1999; Thorrez et al., 2008; Zhao et al., 2021).

In this study, we analysed transcriptome data from 27 libraries of various oil palm tissues to identify genes with low levels of variation in expression. A total of 11 genes with stable expression across all of the tissues tested were identified. From these genes, EgREF_10, annotated as ubiquitin, exhibited the highest level of variation, with highest levels of expression in pollen and floret. Ubiquitins, one of the most commonly used reference genes, are constitutively expressed to maintain cellular function. They are known as housekeeping genes whose protein products are involved in basic cellular processes (Bustin, 2002; Czechowski et al., 2005; Dheda et al., 2004; Hugget et al., 2005). Ubiquitin and small ubiquitin-related modifiers are reported to be involved in post-translational protein modifications by attaching to target proteins to alter their functions (Smalle and Vierstra, 2004). Other functions include protein regulatory activity known as ubiquitination, and ubiquitin-proteasome system, which involves 5% of Arabidopsis proteins (Hellman and Estelle, 2002).

In theory, housekeeping genes are assumed to have a constant level of expression and classical housekeeping genes are frequently used and utilised as reference genes for normalisation in RT-qPCR. A number of frequently used classical housekeeping genes that have been validated as suitable reference genes in many plants include ACTIN, TUBULIN, UBIQUITIN and GAPDH (Garg et al., 2010; Gu et al., 2011; Hu et al., 2009; Maroufi et al., 2010; Nicot et al., 2005). ACTIN, a widely used reference gene, was selected as the most stable reference gene for normalisation of gene expression data in grapevine leaves (Gutha et al., 2010). In papaya, a number of gene expression studies using RT-qPCR had used ACTIN as a reference gene (Hernandez et al., 2007; Kouzaki et al., 2009; Yu et al., 2005; 2008).

TUBULIN was also deemed as an appropriate reference gene in banana (Podevin et al., 2012). This gene was also identified to be the most suitable reference gene for normalisation across various developmental stages of somatic embryos in two representative conifer species, Pinus pinaster and Pinus abies (De Vega-Bartol et al., 2013). As for UBIQUITIN, it was found to have stable expression in A. thaliana and Brachypodium sp. (Czechowski et al., 2005; Hong et al., 2008). It was also considered a suitable reference gene in sugarcane leaf samples (de Andrade et al., 2017) and bioenergy plants (Cheng et al., 2019). GAPDH on the other hand has been used as a reference gene in citrus, Chinese wolfberry, cotton, chickpea, peanut, grapevine, coffee and sugarcane (Barsalobres-Cavallari et al., 2009; de Andrade et al., 2017; Garg et al., 2010; Mafra et al., 2012; Morgante et al., 2011; Reid et al., 2006; Wang et al., 2013a; 2013b).

Among the four classical housekeeping genes, the ranking from the most stable to the least stable gene was: ACTIN > GAPDH > UBIQUITIN > TUBULIN. TUBULIN and UBIQUITIN are two of the five frequently used classical housekeeping genes. However, in our range of samples, these genes performed poorly as well as in tissue culture materials evaluated by Chan et al. (2014). It has been reported that TUBULIN was found poorly ranked in bamboo, peanut, grape, potato and soybean (Chi et al., 2012; Fan et al., 2013; Jan et al., 2008; Nicot et al., 2005; Reid et al., 2006), while UBIQUITIN performed poorly as a reference gene in soybean (Jian et al., 2008) and grape (Reid et al., 2006).

In this study, GAPDH was ranked in the middle amongst the classical housekeeping genes. However, it had unstable expression across papaya fruit samples (Zhu et al., 2012), and was considered the least stably expressed gene during leaf and flower development in petunia (Mallona et al., 2010). It also displayed the biggest variation in leaves and roots of Boehmeria nivea L. (Yu et al., 2020). Based on EST data analysis, this gene was also identified as poorly ranked reference gene in tomato (Coker and Davis, 2003).

In our range of samples, ACTIN was observed to be the most stable gene among all classical housekeeping genes studied. However, two published reference genes and four genes identified from transcriptome data were found to be more stably expressed than ACTIN. Furthermore, expression stability of ACTIN gene family members was observed to differ in peanut in specific conditions (Chi et al., 2012; Morgante et al., 2011; Reddy et al., 2013). These results showed that the transcripts of housekeeping genes can vary in response to experimental conditions and tissue types. Thus, it is important to validate the expression stability of these genes prior to their use in normalisation for RT-qPCR.

Our results showed that three reference genes, EgREF_5, EgREF_7 and EgREF_11, are suitable for normalisation of gene expression data from oil palm tissues. As reported by Szabo et al. (2004), valuable reference genes are expressed in several tissues with minimal variation in transcript level across experimental conditions. Genes with low variation in expression with coefficient variation of normalised relative quantities of less than 0.5
are suitable reference gene candidates (Hellemans et al., 2007). The results suggest that it is possible to identify reliable candidate reference genes from multiple transcriptome datasets. \textit{EgREF\_5} was annotated as deleted in split hand/split foot protein 1A, while the other two are NADH-ubiquinone oxidoreductase 13 kDa-B subunit (\textit{EgREF\_7}) and 40S ribosomal protein S17 (\textit{EgREF\_11}). Expression of these genes were found to be more stable than previously determined reference genes (Chan et al., 2014) and classical housekeeping genes, in the set of diverse oil palm tissues within this study. In recent publications on oil palm reference genes, four genes namely \textit{elf1}, \textit{elf2}, \textit{APT} and \textit{cyc}, were identified from transcriptome datasets by Xia et al. (2014), of which \textit{elf1} and \textit{elf2} were selected as reference genes for cold, drought and salinity treatments. While \textit{APT} and \textit{cyc} were deemed stable reference genes for drought and salinity stress samples, two genes (\textit{PD00380}, \textit{PD00569}) were selected as reference genes for tissue culture samples (Chan et al., 2014). These genes were also determined to be good reference genes for oil palm cultured leaf explants (Ooi et al., 2012) and tissue culture samples (Chan et al., 2010). However, in our range of samples, \textit{PD00380} was observed to be the least stable expressed gene. Nevertheless, \textit{pOP-EA01332} and \textit{PD00569} were observed to have more stable expression and were ranked amongst the top five stably expressed genes. Other oil palm reference genes that have also been identified include \textit{GRAS}, \textit{Glutaredoxin}, \textit{Cyp2} and \textit{SLU7}, where \textit{GRAS} and \textit{Glutaredoxin} are suitable for reproductive tissues, vegetative tissues (\textit{GRAS}, \textit{Cyp2}, \textit{SLU7}) and fruit development (\textit{GRAS}, \textit{Cyp2}) (Yeap et al., 2014). Thus, validation of candidate reference genes is best done under specific experimental conditions and tissue types, as supported by the results of all of these studies.

**CONCLUSION**

This study shows that systematic analysis of transcriptome data can successfully guide the selection of reference genes for gene expression studies in oil palm, resulting in three housekeeping genes that can be used for RT-qPCR experiment on a wide range of mature plant and tissue culture samples. GeNorm statistical algorithm, identified a combination of three genes, \textit{EgREF\_5}, \textit{EgREF\_7} and \textit{EgREF\_11} as suitable reference genes for accurate and reliable normalisation across a diverse range of oil palm samples that consist of leaf, root, endosperm, mesocarp, female flowers, non-embryogenic callus, embryogenic callus, polyploids and shoots from polyploids. Hence, reference genes identified in this study could be used to generate reliable gene expression profile for a wide variety of tissues in oil palm.

**ACKNOWLEDGEMENT**

The authors would like to thank the Director-General of MPOB for permission to publish this article. We also wish to thank the staff of Genomics Laboratory, Tissue Culture Laboratory and Transgenic Technology Laboratory at the Advanced Biotechnology and Breeding Centre, MPOB for their assistance.

**REFERENCES**


Burgis, F; Kilaru, A; Cao, X; Ngando Ebongue, G F; Drira, N; Ohlrogge, J B and Arondel, V (2011).


Chan, K-L; Tatarinova, T V; Rosli, R; Amiruddin, N; Azizi, N; Ab Halim, M A; Sanusi, N S N M; Jayanthi, N; Ponomarenko, P; Triska, M; Solovyev, V; Firdaus-Raith, M; Sambanthamurthi, R; Murphy, D and Low, E T L (2017). Evidence-based gene models for structural and functional annotations of the oil palm genome. *Biol. Direct*, 12: 21.


Chang, C W; Cheng, W C; Chen, C R; Shu, W Y; Tsai, M L; Huang, C L and Hsu, I C (2011). Identification of human housekeeping genes and tissue-selective genes by microarray meta-analysis. *PLoS ONE*, 6: e22859.


de Oliveira, L A; Breton, M C; Bastolla, F M; Camargo, S D; Margis, R; Frazzon, J and Pasquali, G (2012). Reference genes for the normalization of gene expression *Eucalyptus* species. *Plant Cell Physiol.*, 53(2): 405-422.


Dussert, S; Guerin, C; Andersson, M; Joët, T; Tranbarger, T J; Pizot, M; Sarah, G; Omore, A; Durand-Gasselin, T and Morcillo, F (2013). Comparative transcriptome analysis of three oil palm fruit and seed tissues that differ in oil content and fatty acid composition. *Plant Physiol.*, 162: 1337-1358.


Hernandez, M; Cabrera-Ponce, J L; Fragoso, G; Lopez-Casilla, F; Guevara-Garsia, A; Rosas, G; Leon-Ramirez, C; Juaraz, P; Sanchez-Garcia, G; Cervantes, J; Acero, G; Toledo, A; Cruz, C; Bojalil, R; Herrera-Estrella, L and Sciutto, E (2007). A new highly effective anticysticercosis vaccine expressed in transgenic papaya. *Vaccine*, 25: 4252-4260.


Jian, B; Liu, B; Bi, Y; Hou, W; Wu, C and Han, T (2008). Validation of internal control for gene expression study in soybean by quantitative real-time PCR. *BMC Mol. Biol.*, 9: 59.


Li, Y; Liang, X; Zhou, X; Wu, Z; Yuan, L; Wang, Y and Li, Y (2020). Selection of reference genes for qRT-PCR analysis in medicinal plant Glycyrrhiza under abiotic stresses and hormonal treatments. *Plants*, 9: 1441.


Low, E T L; Rosli, R; Jayanthi, N; Mohd-Amin, A H; Azizi, N; Chan, K L; Maqbool, N J; Maclean, P; Brauning, R; McCulloch, A; Moraga, R; Ong-Abdullah, M and Singh, R (2014). Analyses of hypomethylated oil palm gene space. *PLoS ONE*, 9(1): e86728.


Singh, R; Ong-Abdullah, M; Low, E T L; Abdul Manaf, M A; Rosli, R; Nookiah, R; Ooi, C L L; Ooi, S E; Chan, K L; Halim, M A; Azizi, N; Nagappan, J; Bacher, B; Lakey, N; Smith, S W; He, D; Hogan, M; Budiman, M A; Lee, E K; DeSalle, R; Kudrma, D; Goicoechea, J L; Wing, R A; Wilson, R K; Fulton, R S; Ordway, J M; Martienssen, R A and Sambanthamurthi, R (2013). Oil palm genome sequence reveals divergence of infertile species in old and new worlds. *Nature*, 500: 335-339.


Tranbarger, T J; Dussert, S; Joët, T; Argout, X; Summo, M; Champion, A; Cros, D; Omore, A; Nouy, B and Morcillo, F (2011). Regulatory mechanisms underlying oil palm fruit mesocarp maturation, ripening, and functional specialization in lipid and carotenoid metabolism. *Plant Physiol.*, 156: 564-584.


Vandesompele, J; De Preter, K; Pattyn, F; Poppe, B; Van Roy, N; De Paepe, A and Speleman, F (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.*, 3: research0034.1-0034.11.


Xia, W; Mason, A S; Xiao, Y; Liu, Z; Yang, Y; Lei, X; Wu, X; Ma, Z and Peng, M (2014). Analysis of multiple transcriptomes of the African oil palm (*Elaeis guineensis*) to identify reference genes for RT-qPCR. *J. Biotechnol.*, 184: 63-73.


Yu, Q; Hou, S; Feltus, F A; Jones, M R; Murray, J E; Veatch, O; Lemke, C; Saw, J H; Moore, R C; Thimmmapuram, J; Liu, L; Moore, P H; Alam, M; Jiang, J; Paterson, A H and Ming, R (2008). Low X/Y divergence in four pairs of papaya sex-linked genes. *Plant J.*, 53: 124-132.


Yu, Y; Zhang, C; Chen, Y; Bai, Q; Gao, C; Zeng, L; Li, Z; Cheng, Y; Chen, J; Sun, X; Guo, L; Xu, J and Yan, Z (2020). Selection of reference genes for qPCR analyses of gene expression in ramie leaves and roots across eleven abiotic/biotic treatments. *Sci. Rep.*, 9: 20004.

Zhao, Z; Zhou H; Nie, Z; Wang, X; Luo, B; Yi, Z; Li, X; Hu, X and Yang, T (2021). Appropriate reference genes for RT-qPCR normalization in various organs of *Anemone flaccida* Fr. Schmidt at different growing stages. *Genes*, 12: 459.
